

Empirical Bayes estimators for the reproduction parameter of Borel-Tanner distribution

George P. Yanev¹

University of South Florida, Tampa, Florida

ABSTRACT We construct empirical Bayes estimators for the reproduction parameter of Borel-Tanner distribution assuming LINEX loss and prove their asymptotic optimality. Some properties of the estimators regret risk are illustrated through simulations.

AMS Subject Classification 62C10, 62F15, 60J80.

1. INTRODUCTION

The probability mass function of the Borel-Tanner distribution is

$$p(x|\theta, r) = a_r(x)\theta^{x-r}e^{-\theta x}, \quad (x = r, r+1, \dots) \quad (1)$$

where $0 < \theta < 1$, r is a positive integer and $a_r(x) = rx^{x-r-1}/(x-r)!$

Initially (1) was derived as the probability distribution of the number of customers served in a queuing system. It also appears in random trees and branching processes. More specifically, it is the distribution of the total progeny in a Galton-Watson process assuming Poisson reproduction, see Aldous [1] for recent applications. Our interest in estimating θ stems from its role as reproduction number of an epidemic infection modeled by a branching process, see Farrington et al. [2]. We study nonparametric (with respect to the prior) empirical Bayes (NPEB) estimators for θ . The NPEB estimation procedures rely on the assumption for existence of a prior distribution G which, however, is unknown. Consider independent copies $(X_1, \theta_1), \dots, (X_{n+1}, \theta_{n+1})$ of (X, θ) , where θ has a distribution G , and conditional on θ , X has a Borel-Tanner distribution given by (1). The “past” data consist of independent observations x_1, x_2, \dots, x_n obtained with independent realizations $\theta_1, \theta_2, \dots, \theta_n$ of θ , where the X_i s are observable and the θ_i s are not observable. Denote by $\theta_n(x)$ an empirical Bayes estimator for θ based on the “past” data and the “present” observation $x_{n+1} = x$. As Maritz and Lwin [5] point out, an advantage of using NPEB estimators is the minimum assumptions on the class of prior distributions. It turns out that in the case of Borel-Tanner distribution the Bayes rule assuming LINEX loss depends on the prior through the marginals only. This remarkable

¹Address: Mathematics Dept., 4202 E. Fowler Ave, PHY114, University of South Florida, Tampa, FL 33620, USA. E-mail: gyanev@cas.usf.edu.

fact allows us to construct simple NPEB estimators estimating the Bayes rule directly without estimating the prior itself.

Notice that NPEB estimators for θ under weighted squared-error loss are studied in Yanev [7]. In the next section we use the asymmetric LINEX loss function, instead. In Section 3 we prove the estimators asymptotic optimality. The last section contains numerical results concerning the estimators performance measured by their regret risk.

2. EMPIRICAL BAYES ESTIMATION USING LINEX LOSS

In some applications (e.g. surveillance of infectious diseases) the squared-error loss function seems inappropriate in that it assigns the same loss to overestimates as to equal underestimates. A well-known alternative (see Huang et al. [3] and the references therein) is the LINEX loss function defined, for $\gamma_l \leq \gamma \leq \gamma_u$ and $\gamma \neq 0$ by

$$L_\gamma(\hat{\theta}, \theta) = e^{\gamma(\hat{\theta} - \theta)} - \gamma(\hat{\theta} - \theta) - 1, \quad (2)$$

where $\hat{\theta}$ is an estimator for θ . It is clear that the LINEX loss function is convex, asymmetric and for $\gamma > 0$ it increases almost linearly for negative errors and almost exponentially for positive errors. Thus, it penalizes an overestimation more seriously than an underestimation. This is reversed when $\gamma < 0$. For small values of $|\gamma|$ the LINEX loss is close to the squared-error loss. From now on we assume that γ is a positive integer; the case $\gamma < 0$ can be treated similarly.

Based on a single observation, the maximum likelihood estimator $\theta_{MLE}(x)$ for θ is (e.g., Kumar & Consul [4])

$$\theta_{MLE}(x) = \frac{x - r}{x}. \quad (3)$$

Denote by \mathcal{I}_A the indicator of the event A .

Theorem 1 Assume LINEX loss with $\gamma > 0$, integer. A NPEB estimator for θ in (1) is

$$\theta_n(x) = \gamma^{-1} \ln \tau_n(x) \mathcal{I}_{\{\tau_n(x) \in (1, e^\gamma)\}} + (x - r)/x \mathcal{I}_{\{\tau_n(x) \notin (1, e^\gamma)\}}, \quad (4)$$

where

$$\tau_n(x) = \frac{r + \gamma}{r} \left(\frac{x + \gamma}{x} \right)^{x-r-1} \frac{m_n(x|r)}{m_n(x + \gamma|r + \gamma)}$$

and $m_n(z|y)$ is an estimate for the marginal distribution $m_G(z|y) = \int_0^1 p(z|\theta, y) dG(\theta)$.

Proof The Bayesian estimator $\theta_G(x)$ under LINEX loss is (e.g. Huang et al. [3])

$$\theta_G(x) = -\gamma^{-1} \ln E_{G|x} e^{-\gamma\theta}, \quad (5)$$

provided that $E_{G|x} e^{-\gamma\theta} < \infty$, where $E_{G|x}(\cdot)$ is the expectation w.r.t. the posterior. Since

$$\begin{aligned} E_{G|x} e^{-\gamma\theta} &= \frac{1}{m_G(x|r)} \int_0^1 e^{-\gamma\theta} a_r(x) \theta^{x-r} e^{-x\theta} dG(\theta) \\ &= \frac{a_r(x)}{a_{r+\gamma}(x + \gamma)} \frac{m_G(x + \gamma|r + \gamma)}{m_G(x|r)}, \end{aligned}$$

we can write the Bayesian estimator $\theta_G(x)$ from (5) as

$$\begin{aligned}\theta_G(x) &= \gamma^{-1} \ln \left\{ \frac{r+\gamma}{r} \left(\frac{x+\gamma}{x} \right)^{x-r-1} \frac{m_G(x|r)}{m_G(x+\gamma|r+\gamma)} \right\} \\ &= \gamma^{-1} \ln \tau_G(x), \quad \text{say.}\end{aligned}$$

Note that, $\theta_G(x)$ depends upon the prior through the marginal distribution only. Therefore, estimating the marginals, we can construct a NPEB estimator $\theta_n(x)$ for θ as given in (4). \square

One possible form of the estimators $m_n(z|y)$ in Theorem 1 can be obtained as follows. In addition to the current $X_{n+1}(r) = x$, let us have observed n independent pairs

$$(X_1(r), X_1(\gamma)), (X_2(r), X_2(\gamma)), \dots, (X_n(r), X_n(\gamma)), \quad (6)$$

where $X_i(r)$ and $X_i(\gamma)$ are independent and Borel-Tanner distributed with $p(x|\theta_i, r)$ and $p(x|\theta_i, \gamma)$, respectively. It is known (e.g., Kumar & Consul [4]) that $X_i(r) + X_i(\gamma)$ has pmf $p(x|\theta_i, r+\gamma)$. Let $f_n(y|r+\gamma)$ be the number of pairs, such that $X_i(r) + X_i(\gamma) = y$, ($i = 1, \dots, n$). Consistent estimators for the marginals $m_G(x+\gamma|r+\gamma)$ and $m_G(x|r)$ are the relative frequencies

$$m_n(x+\gamma|r+\gamma) = \frac{f_n(x+\gamma|r+\gamma)}{n+1} \quad \text{and} \quad m_n(x|r) = \frac{1+f_n(x|r)}{n+1}. \quad (7)$$

Let us notice here that a NPEB estimator $\tilde{\theta}_n(x)$ for θ under the squared-error loss $L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$ is constructed in Yanev [7] as follows

$$\tilde{\theta}_n(x) = \kappa_n(x) \mathcal{I}_{\{\kappa_n(x) \in (0, 1)\}} + (x-r)/x \mathcal{I}_{\{\kappa_n(x) \notin (0, 1)\}},$$

where

$$\kappa_n(x) = \frac{a_r(x)}{m_n(x)} \sum_{j=0}^{\infty} \frac{(j+1)^{j-1}}{j!} \frac{m_n(x+j+1)}{a_r(x+j+1)},$$

where, as before, $a_r(y) = ry^{y-r-1}/(y-r)!$

3. ASYMPTOTIC OPTIMALITY

The Bayes risk of an estimator $\hat{\theta}$ can be written as

$$R(G, \hat{\theta}) = \int_X \int_{\Theta} L(\hat{\theta}, \theta) p(x|\theta, r) dG(\theta) dx = \sum_{x=r}^{\infty} \int_{\Theta} L(\hat{\theta}, \theta) p(\theta|x) dG(\theta) m_G(x|r),$$

where $p(\theta|x)$ is the posterior distribution. If $R(G, \theta_n|\underline{X}_n)$ is the conditional Bayes risk of the estimator $\theta_n(x)$ given $\underline{X}_n = (X_1, \dots, X_n)$, then $R(G, \theta_n) = E_n\{R(G, \theta_n|\underline{X}_n)\}$ is the (unconditional) Bayes risk of θ_n , where the expectation $E_n(\cdot)$ is taken with respect to \underline{X}_n . The estimator $\theta_n(x)$ is asymptotically optimal for given G if $\lim_{n \rightarrow \infty} R(G, \theta_n) = R(G, \theta_G)$. We shall prove the asymptotic optimality of $\theta_n(x)$.

First, let us find the minimum Bayes risk $R(G, \theta_G)$ attained by the Bayesian estimator $\theta_G(x)$. Since (5) implies $\exp(\gamma\theta_G(x)) \int_0^1 \exp(-\gamma\theta)p(\theta|x)d\theta = 1$, we have

$$\begin{aligned} R(G, \theta_G) &= \sum_{x=r}^{\infty} \left\{ \int_0^1 \left\{ e^{\gamma(\theta_G(x)-\theta)} - \gamma(\theta_G(x) - \theta) - 1 \right\} p(\theta|x) d\theta \right\} m_G(x|r) \\ &= \sum_{x=r}^{\infty} \left\{ e^{\gamma\theta_G(x)} \int_0^1 e^{-\gamma\theta} p(\theta|x) d\theta - \gamma\theta_G(x) + \int_0^1 \gamma\theta p(\theta|x) d\theta - 1 \right\} m_G(x|r) \\ &= \sum_{x=r}^{\infty} \left\{ \int_0^1 \gamma\theta p(\theta|x) d\theta - \gamma\theta_G(x) \right\} m_G(x|r). \end{aligned}$$

Next, using $\exp(\gamma\theta_G(x)) \int_0^1 \exp(-\gamma\theta)p(\theta|x)d\theta = 1$ again, we obtain

$$\begin{aligned} R(G, \theta_n) &= \sum_{x=r}^{\infty} E_n \left\{ e^{\gamma\theta_n(x)} \int_0^1 e^{-\gamma\theta} p(\theta|x) d\theta - \gamma\theta_n(x) + \int_0^1 \gamma\theta p(\theta|x) d\theta - 1 \right\} m_G(x|r) \\ &= \sum_{x=r}^{\infty} E_n \left\{ e^{\gamma(\theta_n(x)-\theta_G(x))} - \gamma(\theta_n(x) - \theta_G(x)) + \int_0^1 \gamma\theta p(\theta|x) d\theta - 1 \right\} m_G(x|r). \end{aligned}$$

Therefore,

$$R(G, \theta_n) - R(G, \theta_G) = \sum_{x=r}^{\infty} E_n \left\{ e^{\gamma(\theta_n(x)-\theta_G(x))} - \gamma(\theta_n(x) - \theta_G(x)) - 1 \right\} m_G(x|r) \quad (8)$$

Let us truncate the Borel–Tanner distribution (1) starting with $r = k$ as follows

$$p^*(x|\theta, k) = \begin{cases} p(x|\theta, k), & \text{if } k \leq x \leq k + N - 1; \\ \sum_{x=k+N}^{\infty} p(x|\theta, k), & \text{if } x = k + N. \end{cases} \quad (9)$$

where N is a positive integer. Denote the truncated marginal by $m_G^*(x) = \int_0^1 p^*(x|\theta, y) dG(\theta)$. Similar to the non-truncated case, if $r \leq x \leq r + N - 1$ then

$$E_{G|x}(e^{-\gamma\theta}) = \frac{a_r(x)}{a_{r+\gamma}(x+\gamma)} \frac{m_G^*(x+\gamma|r+\gamma)}{m_G^*(x|r)} = \frac{a_r(x)}{a_{r+\gamma}(x+\gamma)} \frac{m_G(x+\gamma|r+\gamma)}{m_G(x|r)} = \frac{1}{\tau_G(x)}.$$

If $x = r + N$ then

$$\begin{aligned} E_{G|x}(e^{-\gamma\theta}) &= \frac{1}{m_G^*(r+N|r)} \int_0^1 e^{-\gamma\theta} \sum_{k=r+N}^{\infty} a_r(k) \theta^{k-r} e^{-\theta k} dG(\theta) \\ &= \frac{1}{m_G^*(r+N|r)} \sum_{k=r+N}^{\infty} \frac{a_r(k)}{a_{r+\gamma}(k+\gamma)} m_G^*(k+\gamma|r+\gamma) \\ &= \left(\sum_{k=r+N}^{\infty} m_G(k|r) / \tau_G(k) \right) / \sum_{k=r+N}^{\infty} m_G(k|r). \end{aligned}$$

Let $\tau_G^*(x) = \tau_G(x)$ if $r \leq x \leq r + N - 1$; $= \sum_{k=r+N}^{\infty} m_G(k|r) / \sum_{k=r+N}^{\infty} (m_G(k|r) / \tau_G(k))$ if $x = r + N$. The Bayesian estimator in the truncated case is given by $\theta_G^*(x) = \gamma^{-1} \ln \tau_G^*(x)$. Let us estimate $m_G(x|y)$ by $m_n(x|y)$ as in Theorem 1 and set $\tau_n^*(x) = \tau_n(x)$ if $r \leq x \leq r + N - 1$;

$= \sum_{k=r+N}^{\infty} m_n(k|r) / \sum_{k=r+N}^{\infty} (m_n(k|r) / \tau_n(k))$ if $x = r + N$. We construct a NPEB estimator in the truncated case as follows

$$\theta_n^*(x) = \gamma^{-1} \ln \tau_n^*(x) \mathcal{I}_{\{\tau_n^*(x) \in (1, e^\gamma)\}} + (x - r) / x \mathcal{I}_{\{\tau_n^*(x) \notin (1, e^\gamma)\}}.$$

Now, we are in a position to prove the asymptotic optimality of $\theta_n(x)$.

Theorem 2 Assume prior G with finite first moment. If $m_n(z|y)$ is a consistent estimator for $m_G(z|y)$, then the NPEB estimator $\theta_n(x)$ given by (4) is asymptotically optimal, i.e.,

$$\lim_{n \rightarrow \infty} R(G, \theta_n) = R(G, \theta_G).$$

Proof Since $\theta_G(x)$ is the Bayesian estimator, we have $R(G, \theta_n) > R(G, \theta_G)$ and thus

$$R(G, \theta_n) - R(G, \theta_G) \leq |R(G, \theta_n) - R(G, \theta_n^*)| + |R(G, \theta_n^*) - R(G, \theta_G^*)| + |R(G, \theta_G^*) - R(G, \theta_G)| \quad (10)$$

To prove the theorem it is sufficient to show that the right hand side of (10) has $\lim_{N \rightarrow \infty} \limsup_n$ equals zero, when N is from (9). The truncated analog of (8) leads to

$$|R(G, \theta_n^*) - R(G, \theta_G^*)| = \sum_{x=r}^{r+N} E_n \left\{ e^{\gamma(\theta_n^*(x) - \theta_G^*(x))} - \gamma(\theta_n^*(x) - \theta_G^*(x)) - 1 \right\} m_G(x|r)$$

Since $m_n(z|y)$ is a consistent estimator for $m_G(z|y)$, we have $\lim_{n \rightarrow \infty} \theta_n^*(x) = \theta_G^*(x)$, F^∞ - a.s., where F^∞ is the product measure induced by $X_1, X_2, \dots, X_n, \dots$. Notice that, both θ_n^* and θ_G^* are bounded. Indeed, θ_n^* is bounded by definition and $0 < \theta_G^*(x) = -(1/\gamma) \ln E_{G|x}(e^{-\gamma\theta}) < (1/\gamma) \ln e^\gamma = 1$. Therefore, by the Lebesgue dominated convergence theorem we can pass to the limit inside the expectation in the right hand side above and obtain

$$\lim_{n \rightarrow \infty} |R(G, \theta_n^*) - R(G, \theta_G^*)| = 0. \quad (11)$$

Also, since $p^*(\theta|x, r) = p(\theta|x, r)$, $m_G^*(x) = m_G(x)$ for $r \leq x \leq r + N - 1$, and $m_G^*(r + N) = \sum_{x=r+N}^{\infty} m_G(x)$ it is not difficult to obtain

$$|R(G, \theta_G^*) - R(G, \theta_G)| = \sum_{x=r+N}^{\infty} \left\{ \int_0^1 \gamma \theta (p^*(\theta|r + N) - p(\theta|x)) d\theta - \gamma (\theta_G^*(r + N) - \theta_G(x)) \right\} m_G(x).$$

Since $|p^*(\theta|r + N) - p(\theta|x)| < 1$, $|\theta_G^*(r + N) - \theta_G(x)| < 1$, and $E\theta < \infty$ we have

$$\lim_{N \rightarrow \infty} |R(G, \theta_G^*) - R(G, \theta_G)| = 0. \quad (12)$$

Similar to (12) one can prove that $\lim_{N \rightarrow \infty} |R(G, \theta_n) - R(G, \theta_n^*)| = 0$. This along with (10)-(12) completes the proof. \square

4. NUMERICAL EXAMPLES

Using the notation introduced before (7) we set

$$\tau_n^f(x) = \frac{r + \gamma}{r} \left(\frac{x + \gamma}{x} \right)^{x-r-1} \frac{f_n(x|r)}{f_n(x + \gamma|r + \gamma)} .$$

Let $A = \left\{ \tau_n^f(x) \in (1, e^\gamma) \cap f_n(x + \gamma|r + \gamma) \neq 0 \right\}$ and A^c be its complement. Making use of the relative frequency estimators (7) consider $\theta_n^f(x)$ to be defined by

$$\theta_n^f(x) = \gamma^{-1} \ln \tau_n^f(x) \mathcal{I}_A + (x - r)/x \mathcal{I}_{A^c}.$$

That is, if A occurs, then we estimate θ by $\gamma^{-1} \ln \tau_n^f(x)$; whereas if A^c occurs then we use the MLE (3) for θ instead.

A popular measure of the performance of one estimator $\hat{\theta}(x)$ is its regret risk $S(\hat{\theta}) = R(G, \hat{\theta}) - R(G, \theta_G) > 0$. For our simulation study we take $r = 5$, Uniform $(0.5, 1)$ prior and LINEX loss with $\gamma = 3$. Then the minimum Bayes risk attained by the Bayesian estimator

$$\theta_U(x) = \frac{1}{3} \ln \frac{\int_{0.5}^1 \theta^{x-5} e^{-x\theta} d\theta}{\int_{0.5}^1 \theta^{x-5} e^{-(x+3)\theta} d\theta} ,$$

is $R(U_{(0.5,1)}, \theta_U) = 0.0622$.

In the empirical Bayes scheme (6), let us set $n = 50$. Selecting 50 random values for $\theta_i \sim U_{(0.5,1)}$, $i = 1, 2, \dots, 50$, we generate two sets of 50 branching processes starting with $r = 5$ and $\gamma = 3$ ancestors, respectively and both having $Poisson(\theta_i)$, $i = 1, 2, \dots, 50$ offspring distributions. Notice that the total progeny of each process is a realization of a Borel-Tanner (θ_i, \cdot) random variable. Repeating the above procedure 100 times, we obtain 100 samples of 50 pairs Borel-Tanner observations, $(X_i(5), X_i(3))$, $i = 1, 2, \dots, 50$. Each sample gives us a NPEB estimate $\theta_{50}^f(x)$ with regret risk $S_i(\theta_{50}^f)$, $i = 1, 2, \dots, 100$. We estimate the regret risk $S(\theta_{50}^f)$ with the average $\bar{S}(\theta_{50}^f) = \sum_{i=1}^{100} S_i(\theta_{50}^f)/100$.

The above scheme is repeated with $n = 75$ and $n = 100$. As an illustration, we present in Table 1 results for one sample with $n = 100$. For this particular sample, $S_i(\theta_{100}^f) = 0.0980$, which is less than $S(\theta_{MLE}) = 0.1327$.

x	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$\theta_{100}^f(x)$.46	.69	.92	.65	.58	.51	.55	.53	.62	.96	.61	.69	.16	.72	.79	.75
$\theta_U(x)$.63	.64	.65	.65	.66	.67	.67	.68	.69	.69	.70	.71	.71	.72	.73	.73
$\theta_{MLE}(x)$	0	.16	.28	.38	.44	.50	.55	.58	.62	.64	.67	.69	.71	.72	.74	.75

Table 1: Estimates $\theta_n^f(x)$, $\theta_U(x)$, and $\theta_{MLE}(x)$ for θ from a sample with $n = 100$.

n	$0 \leq x \leq 15$			$0 \leq x \leq 200$		
	$\bar{S}(\theta_n^f)$	$STD(\bar{S}(\theta_n^f))$	$S(\theta_{MLE})$	$\bar{S}(\theta_n^f)$	$STD(\bar{S}(\theta_n^f))$	$S(\theta_{MLE})$
50	0.1211	0.0037	0.1292	0.1397	0.0037	0.1327
75	0.1076	0.0036	0.1292	0.1300	0.0037	0.1327
100	0.1038	0.0033	0.1292	0.1299	0.0036	0.1327

Table 2: Numerical results on regret risks of $\theta_n^f(x)$ and $\theta_{MLE}(x)$.

The numerical results for the regret risks are given in Table 2. Several comments are in place. For small x , (columns 2-4) and $n = 75$ or 100 , the improvement of θ_n^f over θ_{MLE} is substantial. Overall, (columns 5-7), the regret risk of θ_n^f is not higher than that of θ_{MLE} .

Finally, note that the Borel-Tanner distribution (1) has monotone likelihood ratio in x , i.e., $p(x|\theta', r)/p(x|\theta, r)$ is an increasing function of x whenever $0 < \theta < \theta' < 1$. This suggests that the NPEB $\theta_n(x)$ can be improved on by the monotonizing procedure of Van Houwelingen and Stijnen [6].

ACKNOWLEDGEMENTS

I thank R. Gueorguieva for helping me with the simulations done with an Ox version 3.30. I also thank the referee for the valuable comments. This research is partially supported by NFSI-Bulgaria, Grant No. MM-1101/2001.

References

- [1] Aldous, D.J. Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. *Bernoulli* **1999**, 5, 3-48.
- [2] Farrington, C.P.; Kanaan, C.P.; Gay, N.J. Branching process models for surveillance of infectious diseases controlled by mass vaccination. *Biostatistics* **2003**, 4(2), 279-295.
- [3] Huang, S.Y.; Liang, T.C. Empirical Bayes estimation of the truncation parameter with Linex loss. *Statist. Sinica* **1997**, 7, 755-769.
- [4] A. Kumar, A.; Consul, P.C. Minimum variance unbiased estimation for modified power series distribution. *Comm. Statist. A - Theory Methods* **1980**, 9, 1261-1275.
- [5] Maritz, J.S.; Lwin, T. *Empirical Bayes Methods*, 2nd Ed.; Chapman and Hall: London, 1989.
- [6] Van Houwelingen, J.C.; Stijnen, T. Monotone empirical Bayes estimators based on more informative samples. *J. Amer. Statist. Assoc.* **1993**, 88, 1438-1443.
- [7] Yanev, G.P. *Statistical modeling of epidemic disease propagation via branching processes and Bayesian inference*, Dissertation, 2001, University of South Florida.